International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 11157-11163

A Determination of Drinking According to Types of Sentence using Speech Signals

Seong-Geon Bae[1], Won-Hee Lee[2] and Myung-Jin Bae*[3]

[1]School of Software Application, Kangnam University, Gyunggido, Korea.

[2,3]Information and Telecommunication of Department, Soongsil University, 369, Sangdo-ro Dongjak-gu, Seoul, Korea.

[1]Orcid Id: 0000-0003-3252-0062, [2]Orcid Id: 0000-0002-7232-2207, [3]Orcid Id: 0000-0002-7585-0400

**Correspondence Author*

Abstract

In this paper, we investigated the accuracy of negative utterance sentences in order to judge drinking. Drinking makes it difficult to pronounce specific sounds correctly. A number of speech parameters have been studied to identify inaccurate pronunciation features. Generally, the method of grasping the inaccurate pronunciation is analyzed using the change of pitch and formant. Formants have various information such as the individual's uniqueness, clarity, and physical condition, and pronunciation status can also be analyzed. Analysis of the pronunciation through the pronunciation agency may cause difficulty in the analysis due to the pronunciation difference of the individual. On the other hand, analysis of pronunciation through the vocal organs is easy to analyze because consonants or vowels are applied only to the consonant vowel rather than to the pronunciation. Especially when drinking alcohol, the vocal cord is dehydrated due to alcohol. Therefore, it is difficult to elaborate voices or to utter certain sounds. Therefore, in this paper, we have studied the possibility of extracting the alcohol discrimination parameter with the change of the accuracy of the voiced sound according to the vocal sentence vowels. As a result, the longer the sentence, the more pronounced error was obtained and the meaningful result was obtained when the number sentence was spoken.

Keywords: Drinking, spectrum, pronunciation, LPC-coefficients

INTRODUCTION

Of the 131 vessels that were found to have been drunk in the sea during 2015, 90 vessels accounted for 68.7% of the vessels, and the largest percentage. [1] Drinking on the road is easy to crack down on drinking through short distance. However, since it is necessary to stop the movement of vessels such as fishing vessels or cargo vessels from above the sea, it is difficult to control drinking through near places. On the other hand, if it is possible to measure alcohol disturbance at a distance rather than a short distance, it can be applied not only to roads but also to maritime and air traffic. The information that can be obtained from a distance through the state or change of the body is voice. If alcohol is interrupted through voice that can send information to a long distance, it is possible to measure any distance. In order to enforce alcohol abuse by voice over the sea, a wireless device must be provided to connect the sender and the receiver. At sea, all vessels are managed by VTS (Sea Traffic Control System) and provide information services to help the ship in the direction of the ship. When the voice communication is performed through the VTS, the characteristics vary depending on the radio equipment, but the voice signal is distorted during demodulation through the receiving equipment. In the case of a cargo ship or a large ship, it is possible to obtain a distorted voice signal by using a specially-designed radio equipment, while using a relatively low-specification equipment in the case of a small fishing boat. When a signal is demodulated at the time of reception, a phenomenon occurs in which distortion occurs in the voice signal during detection and becomes a limit. If the input voice signal is similar to when the input voice signal is out of the input level and if the possibility of drinking through the distorted voice signal is measured, an error occurs in the judgment rate and it is difficult to intercept the voice signal. It is necessary to extract a parameter that is robust against distortion in order to determine whether or not to drink by using a distorted voice signal on the VTS communication[1][3][4][5].

Therefore, in this paper, we try to determine whether alcohol is strong in the surrounding environment by extracting parameters for determining whether or not to drink according to the type of utterance. In Section 2, we examine the existing method of determining whether or not to drink alcohol. In Chapter 3, we discuss how to extract the alcohol dependency parameter according to the type of vocalization proposed in this paper. In Chapter 4, we conclude the future research direction.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 11157-11163

EXISTING METHODS

Generally speaking, the phonetic organ is classified into two characteristics and analyzed. First, it is analyzed as a formant characteristic having pitch characteristic and negative characteristic having individual characteristic information. Figure 1 shows a picture of the voice generation model[5][6][7][8][9][10].

Fig. 2 and 3 show the difference of formant envelope after treatment according to each LPC coefficient by the analysis method by LPC analysis before and after drinking. The horizontal axis represents frequency and the vertical axis represents energy. The first peak is the first formant and the third peak of the second peak represents the second formant and the third formant, respectively. In the speech signal after drinking, the interval between the vowel and the consonant is characterized by a very significant change in the sound. This means that the accuracy of the vowels connecting the sound and the sound is reduced. When the sound is generated in syllable unit through the saints, it is judged to be a phenomenon because the pronunciation changes due to the change of other vocal organs[3][4][5][6][7].

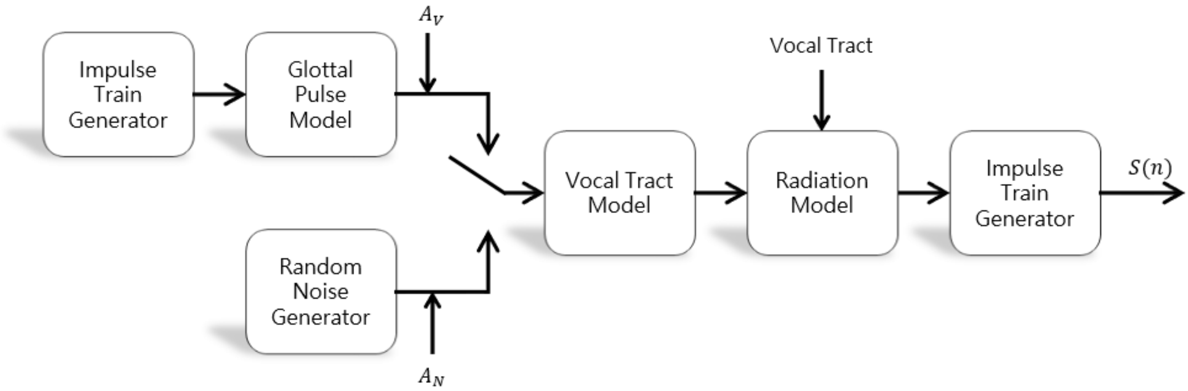


Figure 1: A block of diagram of speech production

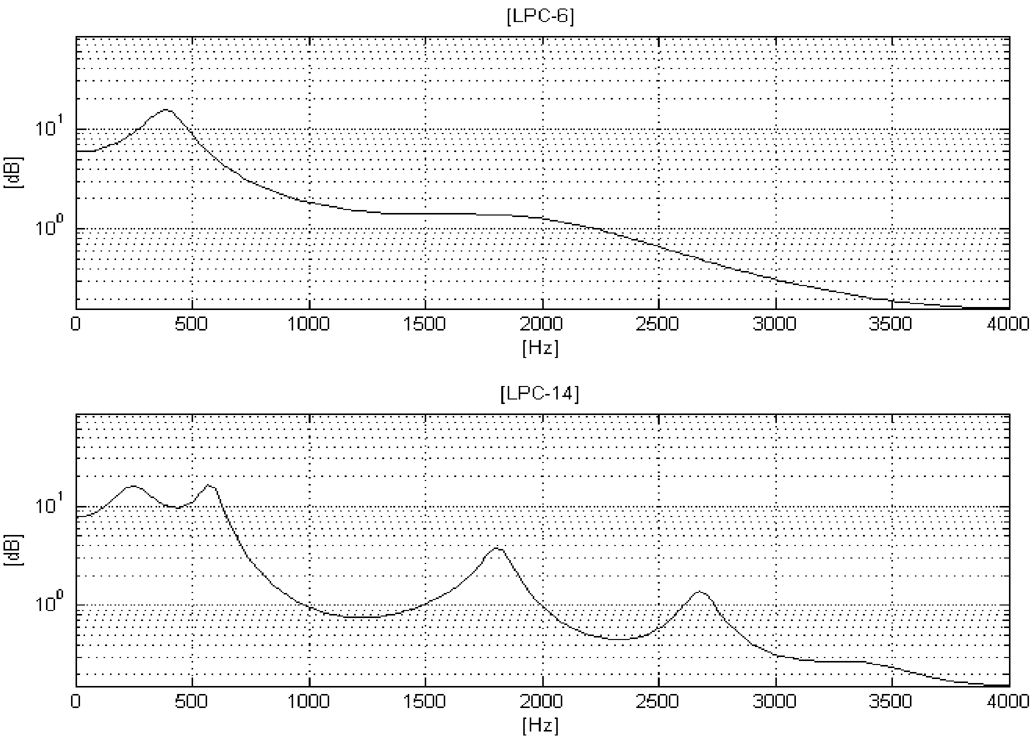


Figure 2: Voice signal formants when the LPC coefficients before drinking are 6, 14

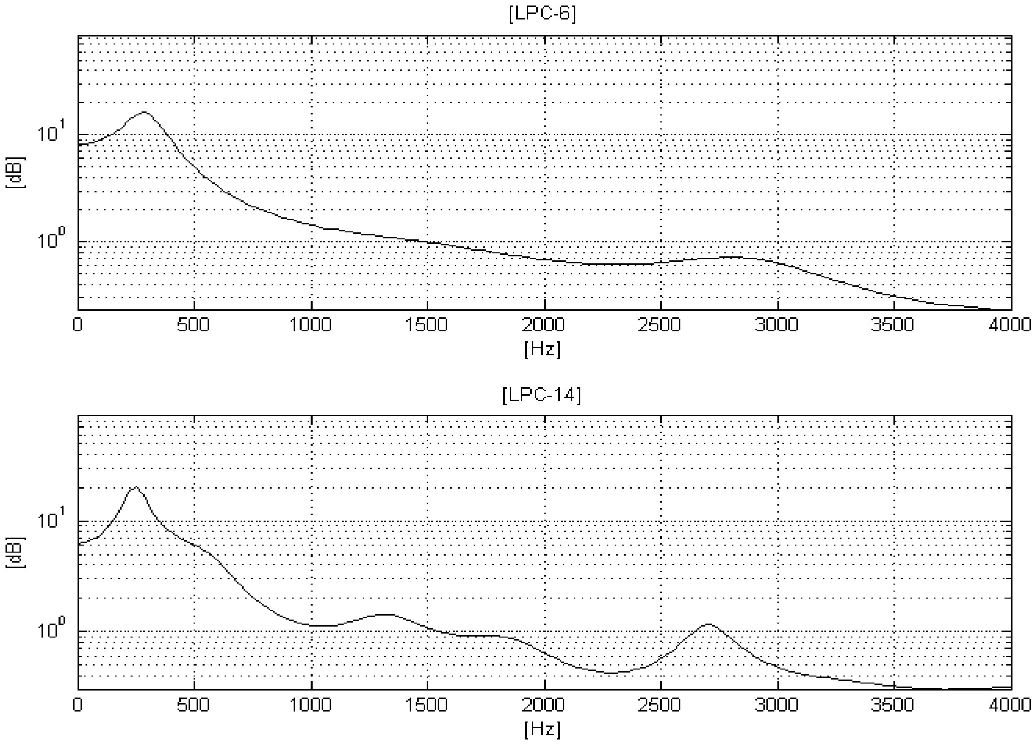


Figure 3: Voice signal formants when the LPC coefficients after drinking are 6, 14

Fig. As can be seen from 2 and 3, there is a difference in formant changes when comparing the voice signals before and after drinking. It can be seen that the change of the third formant and the fourth formant more prominently than the change of the first formant and the second formant. This is because the difference between the 6th and 14th percentiles is significant when calculating formants using LPC before drinking, but not significantly after drinking[3]. Therefore, the envelope for LPC with different orders is obtained in order to measure the formant change amount through the LPC coefficient. Here, 6th and 14th were used for comparison.

3. Proposed Method

Before drinking, voice can be delivered more accurately than voice when drunk. On the other hand, when drinking under the pressure of the vocal cords, the complete opening and closing of the vocal cords may not occur, thereby increasing the minimum pressure required for vocalization and increasing the volume of voice users. The information of the jaw and the tooth can be extracted through the third formant and the fourth formant. The high-frequency domain of the speech signal is useful for analyzing the naturalness and clarity of the sound. In addition, since the third formant and the fourth formant are the formant information in the high frequency region, the parameters can be sufficiently extracted by only the high frequency[1][2][3]. Therefore, in order to investigate the pronunciation accuracy according to sentence types in the speech signal before and after drinking, this paper analyzes through the frequency domain.

Fig. 4 is a block diagram of the procedure for the experiment. In general, since the information of the voice signal is mostly below 4 kHz, the unnecessary components are separated through the low-frequency filter of 3.75 kHz cut-off frequency. The LPF used is defined as having a non-lossy speech signal. Depending on the sentence, fig. 4 method is applied to determine the accuracy of the sound according to the sentence type, and to extract the appropriate sentence when judging drinking. In this paper, the logarithm of the energy of the formants according to each LPC order is obtained and the pronunciation accuracy is obtained because the difference of the formant envelope is pronounced according to the pronunciation sentence type before and after drinking. Equation (1) is shown in Fig. 3 shows the formula of log distance.

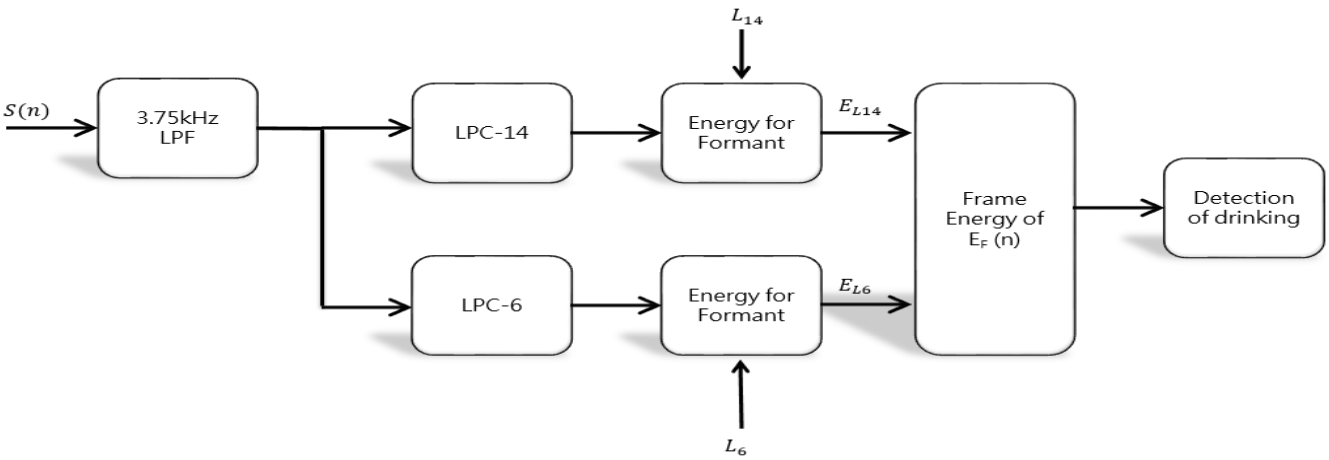
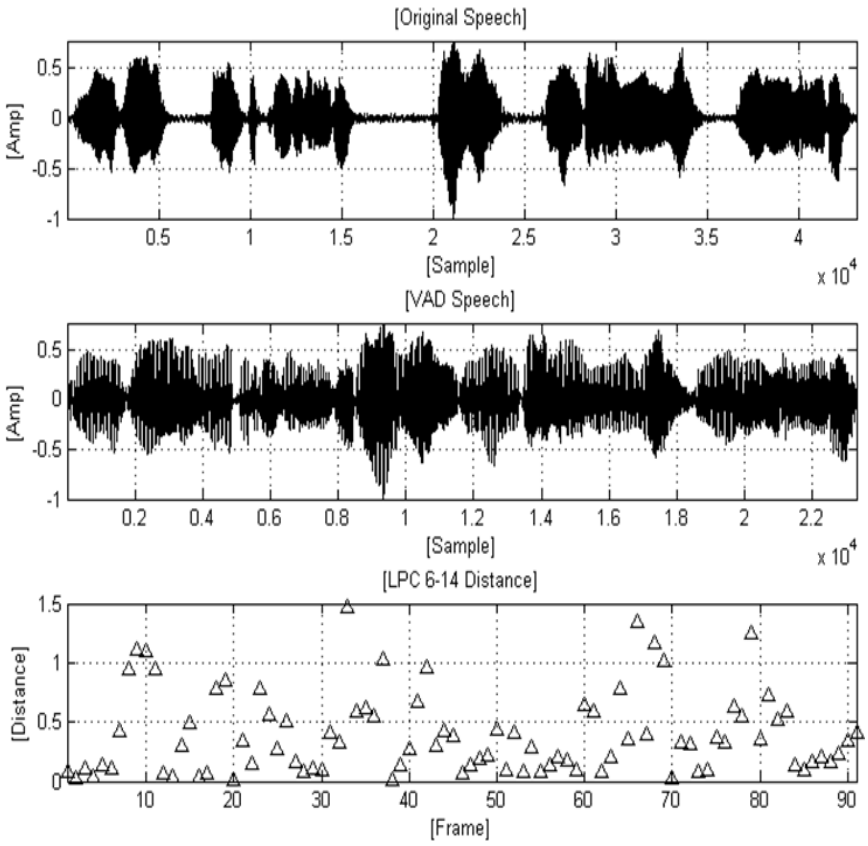


Figure 4: A block of diagram of proposed algorithm

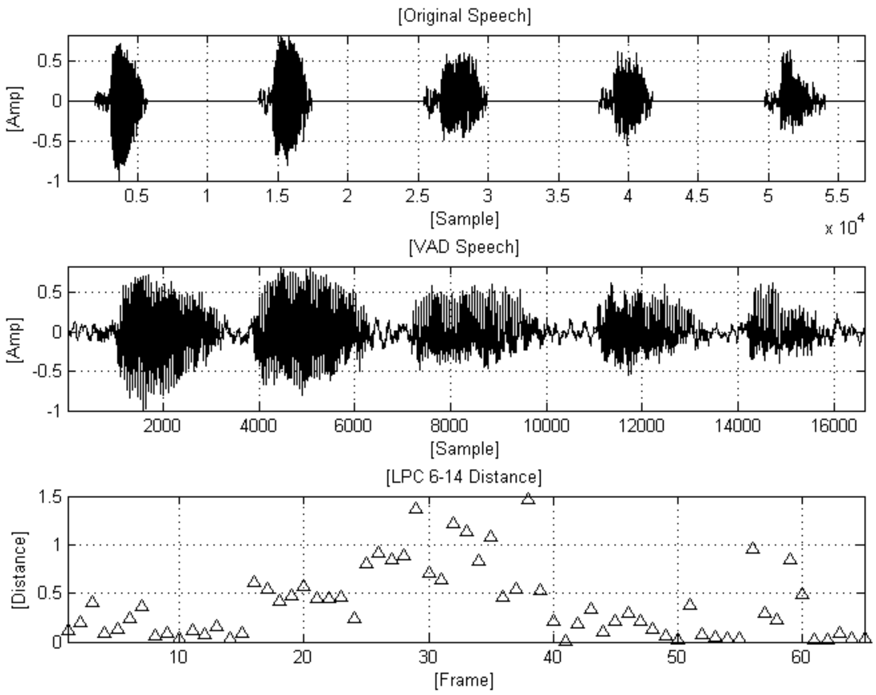
$$E_F(n) = \sqrt{[10 \times \log 10(\frac{L_6 \times E_{L6}}{L_{14} \times E_{L14}})]^2} \quad (1)$$

EXPERIMENTS AND RESULTS

In this paper, the voice before drinking and the voice after drinking were recorded and used as voice samples. The recording environment was recorded at various places such as home, office, laboratory, subway, automobile, bar, and used as a sample. The ages participated in the recording were 20 ~ 60 years old and total 44 voices was analyzed. The amount of alcohol consumed was 20% or more in his own diet. We used the test sentences. Figures 5 to 6 show the proposed method with the speech signal of the sentence.

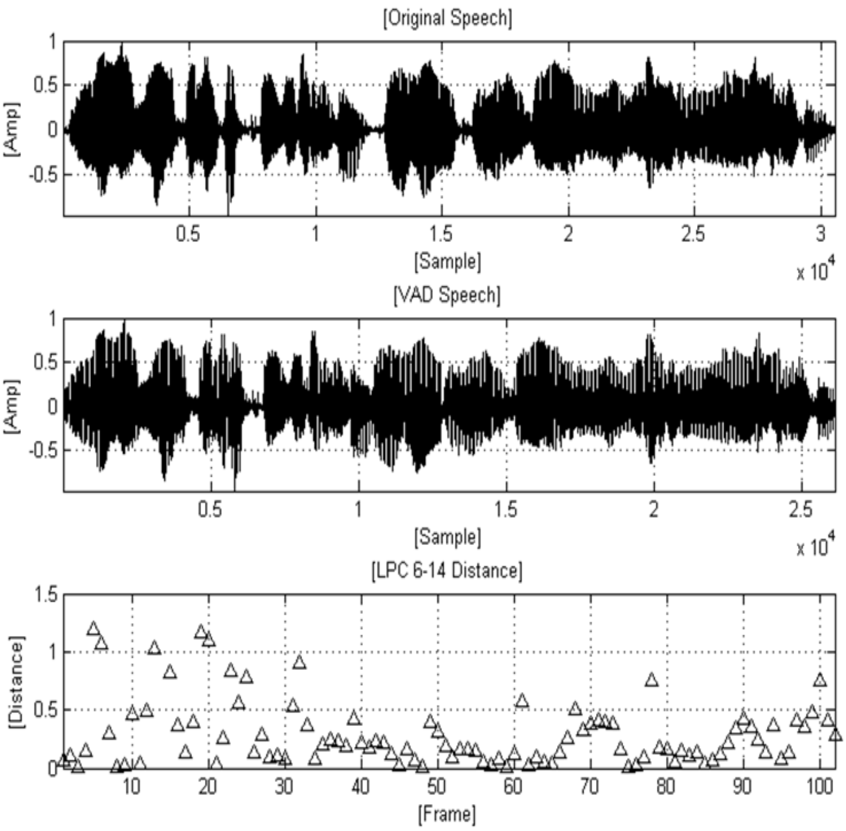


(a)



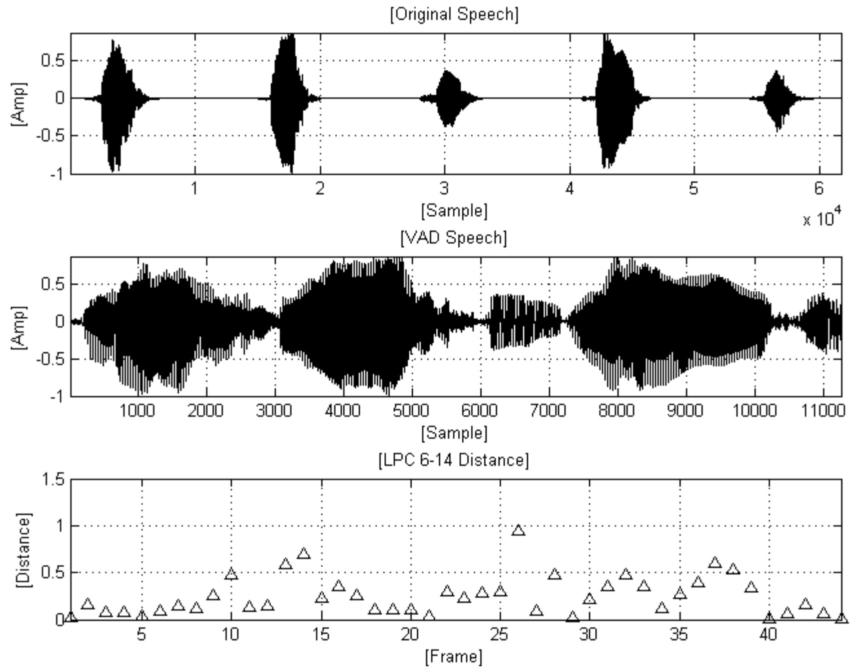
(b)

Figure 4: LPC coefficient difference result before drinking



(a)

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 11157-11163



(b)

Figure 5: LPC coefficient difference result after drinking

As shown in the figure, the whole frame distance of the LPC coefficient in fig 4 before drinking is larger than the coefficient distance in fig. Especially, the difference between "shi" and "sou" is remarkable. Figures show the proposed method with the speech signal of sentence. The phrase "ah-ah-oh" also has a greater overall frame distance after drinking than drinking. However, there were fewer differences than the " shi" sentence. When I saw the phonemes, there was a remarkable result similar to "Shi" in "ie" and "ou". Figures show the proposed method with the speech signal of "one to nine" sentence. "1 to 8," like the other two sentences, the distance difference before drinking is noticeable. In addition, we can see that the difference is larger than the other two sentences. When I saw the phonemes, the results showed that there was a difference between "7" and "8".

CONCLUSION

Drinking accidents occur on the road as much as alcoholic accidents on the road. Drinking accidents occur on the sea, so it is difficult to prevent them beforehand, and if they happen, the cost and damage will increase. If it is possible to find fishing vessels that operate after drunkenness through remote control, it is possible to enforce the crackdown efficiently and reduce the cost and secondary damage. If it is possible to judge the possibility of drinking by using voice data that can transmit a change of a person's body to a remote place, it is possible to sufficiently intercept the communication only in a wireless communication environment. In the case of a cargo ship or a large ship, it is possible to communicate with good sound quality because it uses specially designed equipment and it has less influence on the noise inside the ship. Conversely, in the case of small fishing boats or inexpensive wireless equipment, noise and / or voice signals may be clipped due to the limited input level. Therefore, in this paper, we propose a method to judge the drinking status according to the sentence type in order to be robust to the surrounding environment and to optimize the interception. In the future, this information will be used to gain important information changes in the field of recognition and synthesis

REFERENCES

[1] Seong-Geon Bae, Won-Hee Lee, and Myung-Jin Bae, " A Judgment of Intoxication using Hybrid Analysis with Pitch Coutour Compare in Speech Signal Processing," IJAER, Vol. 12(2017), pp.2342-2346, 2017.

[2] Seong-Geon Bae and Myung-Jin Bae, " A Study on Recovery in Voice Analysis through Vocal Changes before and After Specch Using Speech Signal Processing," IJAER, Vol. 12(2017), pp.5299-5303, 2017.

[3] Seong-Geon Bae, Myung-Jin Bae, "Reducing Errors of Judgement of Intoxication in Overloaded Speech Signal," IJET Vol.8, No.1, pp.219-224, 2016.

[4] Won-Hee Lee, Myung-Sook Kim, and Myung-Jin Bae, “Using valid-frame deviation to judgment of intoxication. Information,” An International Interdisciplinary Journal, Vol. 18 (2015), pp.4131-4136, 2015.

[5] Seong-Geon Bae, Myung-Sook Kim, and Myung-Jin Bae, “Using High Frequency Accentuation in Speech Signals as a New Parameter in Intoxication Judgment. Information,” An International Interdisciplinary Journal, Vol. 17 (2014), pp.6531-6536, 2014.

[6] Seong-Geon Bae, Myung-Jin Bae, “A New Speech Coding using Harmonics Emphasis Filter,” ISAAC 2013, AACL Vol. 1(2013), pp.43-44, 2013.

[7] Geum-Ran Baek, Myung-Jin Bae, “Study on the Judgment of Intoxication State using Speech Analysis ". Vol.352 CCIS, 2012, pp. 277-282, Springer-Verlag. 2012.

[8] Won-Hee Lee, Seong-Geon Bae, and Myung-Jin Bae, "A Study on Improving the Overloaded Speech Waveform to Distinguish Alcohol Intoxication using Spectral Compensation," IJET Vol.7, No.5, pp.235-237.

[9] Chang-Joong Jung, Seong-Geon Bae, Myung-Sook Kim and Myung-Jin Bae, “Speech Sobriety Test Based on Formant Energy Distribution,” IJMUE Vol.8(2013), No.6, pp.209-216, 2013.

[10] S. Marcieli Bellé, Sílvia do Amaral Sartori, Angela Garcia Rossi. “Alcoholism: effects on the cochleo-vestibular apparatus”. Rev Bras Otorrinolaringol. 73(1):116-22. 2007.